World Patent Information
Volume 28, Issue 1, March 2006, Pages 6-13

doi:10.1016/j.wpi.2005.07.007
Cite or Link Using DOI

Automatic classification of patent documents for TRIZ users

Han Tong Loh[a], Cong He[a], and Lixiang Shen[b]

[a]Department of Mechanical Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore

[b]Honeywell Technology Solutions Laboratory, Honeywell Pte Ltd-Singapore, 17 Changi Business Park Central 1, Singapore 486073, Singapore

Available online 1 September 2005.

Abstract

In contrast to traditional inventors, inventors using TRIZ are not only interested in searching for prior art in related fields, but also for the analogous inventions in other fields that have solved the same Technical Contradiction by using the same method. To be useful for TRIZ users, patents are required to be classified by the Contradiction they solved and Inventive Principles they used instead of the fields in which they are involved. Most of the currently available automatic patent classification systems are based on technology-dependent schemes such as the IPC and they cannot satisfy TRIZ users' requirements. In this paper, an automatic patent classification for TRIZ users is proposed and explained in detail. In a preliminary study, patent

documents were collected for 6 out of 40 Inventive Principles, and the proposed automatic classification tested.

Keywords: Patent classification; TRIZ; Inventive principles; Solutions; Contradictions; Automatic classification; Classification precision; Classification recall

Article Outline

1. Introduction

TRIZ is the Russian acronym for the Theory of Inventive Problem Solving developed by Genrich Altshuller in Russia in 1965 [1]. Unlike brainstorming, it is a systematic approach to creativity. Based on his analysis of 40,000 patents, Altshuller recognized that the same fundamental problems (or Contradiction) in one area had been addressed by many inventions in other technological areas. He also found that the same fundamental solutions had been used over and over again. Based upon the 40,000 patents collected, Altshuller summarized 1201 standard engineering problems, named Contradictions[1] [2], and 40 fundamental solutions to these problems, named Inventive Principles [3].

In contrast to traditional approaches to creativity, the inventors using TRIZ are not only interested in searching for inventions in related fields (or prior art), but also analogous problems in other fields that have

previously solved the same Contradiction. By referring to how analogous patents have used the Inventive Principles summarized by Altshuller to solve the same Contradiction, the inventors could be oriented towards the most effective solutions directly, thus saving time and effort. Therefore, to facilitate searching patents for TRIZ users, patents are required to be classified according to the Contradictions and the Inventive Principles.

There are several patent classification schemes currently available. Most of them, such as the International Patent Classification (IPC) [4], classify patents according to the technical fields in which the patents are involved but are inadequate for inventors using TRIZ.

The classification of inventions for TRIZ users has been addressed by some TRIZ software companies such as CREAX [5] and GOLDFIRE [6]. For example, the software, CREAX INNOVATION SUITE, provides some classified examples to explain each Inventive Principle. Although there are numerous merits in the software, e.g. it constructs a systematical process to solve a complex problem, which is quite helpful to generalize the Contradiction(s) addressed by the problem, the number of classified examples is rather low (17 examples on average for each Principle). In addition, they classified the inventions only according to the Inventive Principles and did not consider the Contradictions the patents solved. Therefore, the inventors who are suggested the same Principle are always provided the same examples even when they are solving different Contradictions.

In 2003, Mann and DeWulf [7] presented a new software framework named "Matrix Explorer", which contains a patent database where patent documents were manually classified according to 40 Inventive Principles related to different Contradictions. But the tool "is not available in the public domain due to the sensitivity that some companies may have if they see their intellectual property analyzed for everyone in the world to see" (Mann D, personal communication).

So far, there is no open patent database available with sufficient examples classified by Inventive Principles and Contradiction. One reason is that it is a very time consuming process to manually classify these documents. For example, the classified patent database in Matrix Explorer mentioned earlier is the

result of years of work by 25 full-time patent analysts. These analysts were from various specialty fields and were trained with TRIZ concepts. An important part of their job is to manually label 150,000 US patents with the Contradictions solved by the inventions and the Inventive Principles used [7].

In addition to the huge requirement of manpower and time, the rapid increase of the number of patent applications worldwide makes it harder to classify patents manually. Therefore, there is a need to develop an automated classification system for TRIZ users and TRIZ software developers. In this paper, such an automatic classification system for TRIZ users is proposed and the results of experiments are presented as well.

The rest of this paper is organized as follows. We give an introduction to TRIZ in Section 2, mainly focusing on the 40 Inventive Principles and the Contradiction Table. An example is shown to explain the steps to use TRIZ to invent new solutions. In Section 3, we first explain classical classification schemes of patents and highlight why they are inadequate for TRIZ users. Thereafter the patents collection, representation and processing of our experiment are introduced. Section 4 presents the results and analysis of our experiments. Possible further research is proposed in Section 5.

2. TRIZ approach to creativity

TRIZ, an innovation methodology, provides a systematic process to define and solve any given problems. It is different from the traditional trial and error approach which mainly relies on brainstorming and becomes unreliable with increased complexity of the inventive problem. Fig. 1 shows the difference between the traditional approach and the TRIZ approach to creativity.

Full-size image (25K)

Fig. 1. Traditional approach to creativity (a) and TRIZ approach to creativity (b) [9].

As we can see, the traditional approach jumps from "my problem" to "my solution" directly, which is restricted by the inventors' personal knowledge. Each researcher has his own specialty and favorite directions for investigation, known as psychological inertia, which "influences researchers to move in the same direction as they have on successful project searches in the past". "This situation resembles a laboratory rat that traces only one path in the maze of world knowledge [1]". Using TRIZ, however, inventors firstly map their specific problems to "analogous standard problems" and then get their solution by referring to the most useful solutions (or Principles) to solve analogous problems, which might come from other technological fields. Innovation by TRIZ is no longer a random process restricted by inventors' psychological inertia. Instead, TRIZ directs the inventors to "access to the knowledge and experiences of the world's finest inventive minds" [5].

With the continuous effort of introducing TRIZ, more and more people have been impressed by the power of this innovative approach to creativity. It has been used by many famous companies such as Ford, Motorola, Siemens, Phillips [8].

2.1. Forty inventive principles and the contradiction table

After initially reviewing over 200,000 of the world's most successful patents, Altshuller focused on 40,000 of them as representative of inventive problems which are the ones containing at least one Contradiction, where improvement of one parameter detracts from another parameter. Based on research upon these patents, TRIZ were developed. Two of the key findings of TRIZ research are 40 Inventive Principles and the Contradiction Table [5].

During his study, Altshuller found that more than 90% of the engineering problems had been solved before [9]: the same fundamental problems (or Contradictions) in one area had been addressed by many inventions in other technological areas and the same fundamental solutions had been used over and over again. Based on the analysis of 40,000 patents, which Altshuller abstracted to 40 Inventive Principles, he then constructed the Contradiction Table to resolve over 1200 Contradictions between pairs of 39 standard engineering parameters [2].

Using TRIZ, the inventors firstly define the specific problems they want to solve, identify the corresponding Contradictions and then look up the Principles from the Contradiction Table. Since the suggested Principles were summarized from numerous inventions that had successfully solved the corresponding Contradictions, they are highly likely to be useful in solving the current problems.

2.2. TRIZ steps to solve problems

To illustrate the TRIZ approach to creativity, an example about "designing of beverage cans" is shown as follows [10]:

Step 1: Identify a problem.

When designing beverage cans, the walls of cans should be as thin as possible. However the cans cannot support a large stacking load if the walls are too thin. The usual engineering solution is to compromise by a trade-off between the thickness requirement and the strength requirement. The ideal result is to solve this Contradiction without trade-off.

Step 2: Formulate this problem using "TRIZ language".

At this step, the specific problem of designing a can could be generalized to an abstract engineering problem: to solve the Contradiction between "Parameter 4, length of a nonmoving object"[2] and "Parameter 11, stress".

Step 3: Search for previously analogous solutions and adapt to "my solution".

From the Contradiction Table, Principle 1, 14 and 35 are suggested to solve the Contradiction between "length" and "stress". Using Principle 1, for this example, the wall of the can could be corrugated or wavy with a lot of "little walls" as illustrated in Fig. 2. With this corrugated wall, the edge strength of the wall would be increased yet allowing a thinner material to be used.

Full-size image (3K)

Fig. 2. Cross-section of corrugated can wall (the improved design using Principle 1) [9].

In Step 3, the Contradiction Table suggests some Principles which are supposed to be the most useful to solve the Contradiction concerned. These suggested Principles provide useful hints to direct the inventors to possible solutions. Yet it is more helpful if inventors are provided specific examples about how former inventors have used these Principles to solve the Contradiction. By doing so, inventors could find inspiration more easily. Therefore, classified patents according to Inventive Principles associated with different Contradictions can provide quite helpful references to the inventors who are seeking for solutions to their problem. As the Principles themselves are rather abstract and there are different ways to use each Principle, giving examples of patents which are the closest match for the inventors can be an added advantage. That is why classifying patents according to Inventive Principles is helpful. To be able to do so automatically or semi-automatically will save much time and efforts. In the next section, our experiment of automatic patent classification for TRIZ users is shown and analyzed.

3. Automatic classification of patent documents

In this section, we will first introduce currently popular patent classification schemes, based on which some research on automatic classification has been reported, and then we analyze why these works are inadequate for our purpose. Thereafter, our experiments of proposed automatic classification for TRIZ users will be shown in detail.

3.1. Classification schemes

To search for inventions in a related field (or prior art), many patent classification schemes, such as IPC and US Patent Codes, have been developed. Research on automatic patent classification utilizing these

schemes has been reported by some researchers [11]. Larkey [12] and [13] created a system to automatically classify US patents into US Patent Codes. Krier and Zaccà [14] reported their research on "automatic categorization applications at the European patent office". Fall et al. [11] published their results of automatic classification in the International Patent Classification.

However, the classification schemes used by these researchers are based on the application fields involved in the inventions. For example, "the IPC divides all technological fields into sections designated by one of the capital letters A to H [11]":

A

Human necessities

B

Performing operations, transporting

C

Chemistry, metallurgy

D

Textiles, paper

E

Fixed constructions

F

Mechanical engineering, lighting, heating, weapons, blasting

G

Physics

H

Electricity

Automatic classification built upon this kind of field-dependent schemes is helpful to search the prior art for traditional inventors. However, it is inadequate for TRIZ users since TRIZ users are interested in previous patents that have solved the same Contradiction and used the same Inventive Principles, which may come from different fields. A patent database with patents which are classified according to Inventive Principles combined with Contradiction

provides a broader view for inventors using TRIZ, by helping them find possible inspiration from a field that may be totally different from theirs.

3.2. Automatic classification of patent documents for TRIZ users

We have built up a small patent database of US patents. Currently, the patents are classified only according to Inventive Principles they used, without considering the Contradictions they solved.[3]

At this stage, two simplifications are made: (1) classification is limited to single-label specifications, assuming that each patent only solves one Contradiction by using one Principle, i.e. each document is only associated with one class; (2) the assumption of class balance is made, i.e. the prototype database has even number of documents in each class. Based on these two simplifications and the collected documents, some preliminary work on automatic patent classification for TRIZ users has been done.

3.2.1. Document collection for 6 Principles

As mentioned earlier, currently there is no open patent database available with sufficient examples classified by Principles. To build a data set for our experiment, we have to manually collect patent documents mainly referring to the brief description of the classified examples in [3] and [15]. From 40 Principles, we selected 6 Principles which are the ones with the most available patent documents by referring to the examples in [3] and [15]. The selected 6 Principles are listed as following: (please refer to [2] for description in detail)

• Principle 1 segmentation,

• Principle 4 asymmetrical,

• Principle 14 spheroidality,

• Principle 17 moving to a new dimension,

• Principle 18 mechanical vibration,

• Principle 35 change parameter.

The patent documents in our experiment are all collected from USPTO (United States Patent and Trademark Office) Patent Full-Text and Image

Database [16], filed from 1976 to the present. For class balance, an equal number of documents for each Principle are collected (25 documents per class). In total the current set contains 150 patent documents. In addition, since single-label is assumed in the experiments, the collected documents are highly distinct from one another, which means that if the patent uses one Principle, it does not use any of the other five.

3.2.2. Representation of the documents

Each patent document in the database comprises several parts. Some parts provide numerical information such as patent number, date of application and figure number. Some parts contain specific pieces of text information, such as names of authors and patent examiners. Other parts are narrative text providing information regarding the patent and are given under the headings [12]:

- Title,
- Abstract,
- Background summary,
- Detailed description,
- Claims.

There are many ways, in previous patent retrieval applications, to represent the whole patent documents. For example, Liang et al. [17] believes that the human generated abstracts of patent documents are very precise and are regarded as the most important part. He supposed that "the abstracts are equivalent to their documents" and used the abstracts to represent the whole documents in his experiment. Fall et al. [11] separately used (a) the titles, (b) the claims sections, and (c) the first 300 words of documents to represent the documents and found that the best performance is achieved using the last representation.

When manually classifying the documents, we found that usually the abstracts and summaries provided enough semantic information to determine the Inventive Principles that the patents used. Therefore in our experiment, we represent the documents in two different ways: (a) their titles and abstracts, (b) their titles, abstracts and summaries[4] and then compare

the results achieved by each representation.

3.2.3. Processing of the documents

Document indexing is performed at word level by calculating the word frequencies in each document (term frequency).[5] After removing stop words [18] in each document, word stemming [19] is performed. We then choose three commonly known metrics to select features: Information Gain (IG), Chi-Square (CHI) and Document Frequency (DF, setting the threshold at 2, 3 and 5 separately) [20]. The number of words after processing at each step is listed in Table 1.

Table 1.

The total number of words before preprocessing and after each step of processing

		Title +
Before preprocessing		19,78C
After stopword removing		10,808
Number of different words		2972
After stemming and removing repetition		1928
After feature selection by IG/CHI		19
After feature selection by DF	Threshold = 2	889
	Threshold = 3	607
	Threshold = 5	340

The stop-word list from [18] is modified in this experiment based on the analysis of Inventive Principles. Some adverbs and pronouns, which are on the stop-word list for general usage and removed as irrelevant information, might provide important information about Inventive Principles. e.g. the words "first … second …" are usually used in the patents using Principle 1, segmentation. The new stop-word list is modified mainly considering the 6 Principles concerned in this experiment.

4. Results and analysis

Table 2 shows the results of automatic classification of the 150 collected patent documents based on (a) their titles and abstracts and (b) titles, abstracts and

summaries. Evaluating using the test mode of 10-fold cross-validation [21], we compared the classification accuracy of different classifiers: *k*-nearest neighbor (*k*NN), decision tree (DT), support vector machine (SVM), and Naı¨ve Bayes (NB) [22] with different feature selection metrics: IG, CHI and DF by WEKA [23]. We found that after "summary" was added to represent documents, the accuracy was decreased in most cases (Table 2), but the number of words to be processed has been increased by several fold (Table 1). In terms of feature selection techniques, IG/CHI performed better than DF in all cases. Among four classifiers used in our experiment, DT performed the best when the dimension of features (i.e. the number of selected features) is low (using IG to select features); NB performed the best, SVM the second best and DT performed the worse in most cases when the dimension of features is high (using DF to select features).

Table 2.

Accuracy of automatic classification based on (a) titles and abstracts (T + A) and (b) titles, abstracts and summaries (T + A + S)[a]

Representation of documents		
Feature selection by IG		T + A
		T + A + S
Feature selection by DF	Threshold = 2	T + A
	Threshold = 3	
	Threshold = 5	
	Threshold = 2	T + A + S
	Threshold = 3	
	Threshold = 5	

[a] In our experiment, the features that are selected by IG and CHI are almost the same. Therefore we only present the performance of feature selection using IG and DF.

As shown in Table 2, the highest accuracy (66.7%) was achieved when representing documents by their "title + abstract + summary", selecting features by IG and using DT as the classifier. It is also shown that the accuracy is 66%, only marginally lower than the

highest accuracy (66.7%), when representing documents by their titles and abstracts, selecting the features by IG and using DT or SVM as the classifiers.

Table 3 presents the confusion matrix and Table 4 displays classification precision, recall and *F*-measure (a combination between precision and recall, defined as $\frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}}$) for each Principle under this situation. From Table 3 and Table 4, we can see that classification result is quite good for some Principles, but poor for some others. Among the 6 Principles, the best performance was achieved for Principle 18, "mechanical vibration", with the highest precision (0.957), the highest recall (0.88) and the highest *F*-measure (0.917) as a result. However, the performance for Principle 17, "moving into a new dimension", was poor mainly because of the low precision (0.378). As shown in Table 3, quite a few patents from other classes, e.g. Principle 1 and 14, were misclassified as the ones using Principle 17, which leads to the low precision.

Table 3.

The confusion matrix under the situation where the highest accuracy (66.7%) was achieved

P1	P4	P14	P17	P18	P35	Classified as
14	1	2	7	0	1	P1
2	14	2	6	0	1	P4
0	1	16	7	0	1	P14
0	3	2	17	1	2	P17
0	0	0	3	22	0	P18
1	2	0	5	0	17	P35

Table 4.

Classification precision, recall and *F*-measure for each Principle under the situation where the highest accuracy (66.7%) was achieved

Principle	Precision	Recall	*F*-measure
P1	0.824	0.56	0.667
P4	0.667	0.56	0.609
P14	0.727	0.64	0.681
P17	0.378	0.68	0.486
P18	0.957	0.88	0.917
P35	0.773	0.68	0.723

Table 4 shows that the differing performance of different Principles. The main reason for the variance is that the documents in some Principles contain obvious and sufficient text information to be differentiated. For example, most inventions among patent documents utilizing Principle 18: e.g. "Electric carving knife with vibrating blades", "Distribute powder with vibration" and "Quartz crystal oscillations drive high accuracy clocks" [3], contain at least one of these words: "vibration", "ultrasonic" and "oscillation", etc. As a result, it is easy to discriminate patent documents using this kind of Principle since most of them contain a clear and similar description literally. However, for other Principles, few documents share similar text information. For example, the patents using Principle 17 "moving to a new dimension" [3]: e.g. "Five-axis cutting tool", "Infrared computer mouse" and "Cassette with 6 CD" have text information that are very different and it is quite hard to cluster them together using text matching alone. How to increase the performance of this kind of "challenging" Principle needs to be further explored.

5. Conclusion and future research

Classified patents according to Inventive Principles and Contradiction are required for TRIZ users. Currently, however, we are lacking open databases with sufficient classified patents of this kind partly because of the huge manpower requirement of manual classification. With a wider application of TRIZ and enormous increase of patents worldwide, there is an urgent need to automatically classify patents for TRIZ users. The main purpose of this paper is to present a new topic of automatic patent categorization in TRIZ categories. In this paper, an automatic patent classification according to Inventive Principles is presented, using 6 of the 40 Principles for a start. For each Principle, 25 patent documents were collected from USPTO Patent Full-Text and

Image Database, represented in two different ways and indexed at word level. The performance of automatic classification was compared based on different feature selection metrics (IG, CHI and DF) and different classifiers (*k*NN, DT, NB and SVM) using the WEKA software. In our experiment, the classification performance is different for different Principles. It was also shown that classification accuracy decreased in most cases after adding "summary" to represent the documents and that IG and CHI performed better than DF. In terms of classification algorithms, DT performed the best if the number of features is low, while NB performed the best if the number of features is high.

The main purpose of this paper is to present a new topic of automatic patent categorization in TRIZ categories. The preliminary experiment shown in the paper is just a start of research on the new topic and has raised several questions for future research. Firstly, more work is needed to analyze the classification performance of the rest of the Principles and to take into account Contradictions-related classification as well. Secondly, the issues of multi-label classification and class imbalance should be addressed. In practice, one patent often involves more than one Contradiction and uses several Inventive Principles. Also, the usage frequency of different Principles varies widely and the class distribution of patents is uneven.

Acknowledgments

The authors would like to thank Mr. Sim Song Wee for collection of patent documents, Mr. Zhang Jun and Mr. Ivan Yap for helpful discussions and suggestions.

References

[1] J. Terninko, A. Zusman and B. Zlotin, Systematic innovation: an introduction to TRIZ, St. Lucie Press (1998).

[2] Contradiction Table. Available from: http://security.westserver.net/vaner/triz4u/triz/download/free/CTTable.pdf.

[3] K. Tate and E. Domb, 40 Inventive Principles with examples, *TRIZ J* (1997) (July).

[4] M. Makarov, The seventh edition of the IPC, *World Patent Inform* **22** (2000), pp. 53–58. **Article** | PDF

(65 K) | View Record in Scopus | Cited By in Scopus (1)

[5] The trial version of CREAX INNOVATION SUITE 3.1. Available from: http://www.creax.com/trialVersion/evaluation.html.

[6] Goldfire http://gfi.goldfire.com/, a presentation about introduction to this software is given in a Singapore company: http://www.cadit.com.sg/.

[7] Mann D, DeWulf S. Updating the Contradiction matrix. In: 5th Annual International Conference of Altshuller Institute for TRIZ Studies, Philadelphia, USA, 2003.

[8] What is TRIZ? Available from: http://triz-journal.com/whatistriz_orig.htm.

[9] Theory of inventive problem Solving (TRIZ). Available from: http://www.mazur.net/triz/.

[10] TRIZ. Available from: http://www.massey.ac.nz/~odiegel/trizworks/TRIZ.doc.

[11] Fall CJ, Torcsvari A, Benzineb K, Karetka G. Automated categorization in the international patent classification. In: ACM SIGIR forum April 2003, vol. 37 (1).

[12] L.S. Larkey, A patent search and classification system, *Digital Libraries 99—The Fourth ACM Conference on Digital Libraries, Berkeley, CA, 11–14 August*, ACM Press, New York (1999), pp. 79–87.

[13] Larkey LS. Some issues in the automatic classification of US Patents. In: Learning for Text Categorization. Papers from the 1998 Workshop. AAAI Press, Technical Report WS-98-05, p. 87–90.

[14] M. Krier and F. Zaccà, Automatic categorization applications at the European patent office, *World Patent Inform* **24** (2002), pp. 187–196. **Article** | PDF (771 K) | View Record in Scopus | Cited By in Scopus (10)

[15] 40 Invention Principles with examples. Available from: http://www.oxfordcreativity.co.uk/.

[16] USPTO patent full-text and image database: http://patft.uspto.gov/netahtml/search-bool.html.

[17] Liang C, Naoyuki T, Hisahiro A. A patent document retrieval system addressing both semantic and syntactic properties. In: ACL Workshop on Patent Corpus Processing, 2003.

[18] Onix Text Retrieval Toolkit. Available from: http://www.lextek.com/manuals/onix/stopwords1.html.

[19] Porter Stemming Algorithm. Available from: http://www.tartarus.org/~martin/PorterStemmer/.

[20] Yang Y, Pedersen JO. A comparative study on feature selection in text categorization. In: Proceedings of the 14th International Conference on Machine Learning ICML97, 1997, p. 412–20.

[21] TenFoldCrossValidation. Available from: http://wiki.apache.org/spamassassin/TenFoldCrossValidation.

[22] Yang Y, Liu X. A re-examination of text categorization methods. In: SIGIR-99, 1999.

[23] WEKA. Available from: http://www.cs.waikato.ac.nz/ml/weka/.

Additional bibliography

J. Rampelmann, Classification and the future of the IPC—the EPO view, *World Patent Inform* **21** (1999) (3), pp. 183–190.

M. Makarov, The process of reforming the international patent classification, *World Patent Inform* **26** (2004) (2), pp. 137–141.

M. Orloff, Inventive thinking through TRIZ: a practical guide, Springer Verlag (2003) ISBN 3-540-44018-6.

Altshuller GS. 40 Principles: TRIZ keys to technical innovation, translated and edited by Shulyak L, Rodman S, Technical Innovation Center, Worcester, MA, 1997.

TRIZ. Available from: http://www.mv.com/ipusers/rm/TRIZ.htm.

Zhang J. Systematic innovation in service design through TRIZ. Master thesis, 2004.

H. Smith, Automation of patent classification, *World Patent Inform* **24** (2002), pp. 269–271.

S. Adams, Comparing the IPC and the US classification systems for the patent searcher, *World Patent Inform* **23** (2001) (1), pp. 15–23.

A. Webb, TRIZ: An inventive approach to invention, *CIPA J* (2002) (November), pp. 541–547.

Article | PDF (86 K) **Article** | PDF (238 K) | View Record in Scopus | Cited By in Scopus (7) **Article** | PDF (54 K) | View Record in Scopus | Cited By in Scopus (4) **Article** | PDF (117 K) | View Record in Scopus | Cited By in Scopus (3)

Corresponding author. Tel.: +65 68742553; fax: +65 67791459.

[1] In TRIZ, two kinds of Contradictions are defined: Technical Contradictions and Physical Contradictions. In this paper only Technical Contradictions are mentioned. To be concise, the word "Contradictions" in this paper is used to denote "Technical Contradiction".

[2] The parameter, "Length", refers to any linear dimension including diameter, width and height, etc.

[3] In this paper, we classify the documents only according to the Principles due to the limitation of time. Since there are over 1200 Contradictions, more time are required to collect patents classified by each of the Contradiction. The main purpose of this paper is to present a new topic of automatic categorization in TRIZ categories. Automatic classification according to Principles is tested at the first stage, followed by Contradiction-related classification in the future.

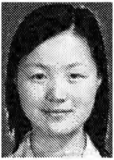
[4] For several documents that do not contain summaries, the descriptions are used instead.

[5] We also tried to use tf * idf (term frequency * inverse document frequency) to weight the words, but found that there was no obvious difference from using term frequency.

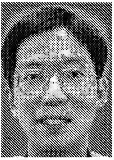
Vitae

Loh Han Tong received his Bachelor in Engineering from the University of Adelaide, his Master of Engineering from the National University of Singapore (NUS) and his Master of Science and PhD in Mechanical Engineering from the University of Michigan. He is an Associate

Professor and Deputy Head in the Department of Mechanical Engineering at NUS. He is also a Fellow of the Singapore-MIT Alliance, which is an innovative engineering education and research collaboration between MIT, NUS and the Nanyang Technological University, to promote global education and research in engineering. His research interests include data mining, rapid prototyping, robust design and computer aided design.



He Cong is currently pursuing a PhD in the Department of Mechanical Engineering from the National University of Singapore, focusing on text mining and TRIZ. She has received her Bachelor in Engineering from the Huazhong University of Science and Technology in China in July, 2002.



Shen Lixiang is a Senior Research Engineer with Technology Solutions Lab, Honeywell Pte Ltd, Singapore. Prior to joining Honeywell, he was an Engineer (Data Mining) in Design Technology Institute Ltd. He was a key member of the team that won Honourable Mention in KDD Cup 2002 (Task 1). He earned his PhD degree from National University of Singapore in 2002 on data mining. His current research interests are in applications of data/text mining to Abnormal Situation Management (ASM).

World Patent Information
Volume 28, Issue 1, March 2006, Pages 6-13